3. TECHNICAL APPROACH

The objective of this project is to ensure that a consistent and reliable set of CCOS meteorological and air quality data is ready for use by data analysts and modelers. To meet these objectives, we will complete the following tasks as outlined in the RFP:

- Tasks 1 and 2 will provide information on which data are available and which data are missing. As part of Task 2, we will take remedial action to obtain the missing data, when feasible. The acquired missing data will be made available to the CCAQS data manager for inclusion in the database.
- In parallel with Tasks 1 and 2, and as part of Task 5, we will contact data contractors and data users to determine whether the most recently QC'd data are in the CCAQS database. We will take remedial action to obtain the most recent versions of the data and associated QC flags if they are not in the database. These acquired data will be made available to the CCAQS data manager. The result at this point will be an updated data set containing previously missing data that were resubmitted by contractors and reflect the most recent QC efforts performed by data contractors and users.
- The new data set will then be subjected to gross outlier checks (Task 3) and metadata checks (Task 4). After Tasks 3 and 4 are completed, the data set will contain additional QC codes that reflect issues found as part of these tasks.
- In parallel with Tasks 1 through 5, as part of Task 6 we will perform a subjective QC of high priority "fast-track" data sets that are needed for current analysis and modeling efforts.
- In addition, for Task 6 we will also QC other priority data sets identified by the TC once Task 1 through 5 are complete.
- Upon completion of all QA/QC undertaken in this effort, we will document in a final report (Task 7) the results and findings regarding the quality and availability of CCOS data in the CCAQS database.

3.1 TASK 0. PREPARE WORK PLAN

In the first task, we will prepare a work plan that describes the details of the work to be performed in Tasks 1 through 7. The work plan will expand upon the proposal and address questions raised by the TC. Once the work plan has been approved, work on the following tasks can commence.

3.2 TASK 1. COMPILE INVENTORY OF CCOS DATA

The objective of this task is to prepare a complete inventory of the CCOS field data sets that are expected to reside in the CCAQS database. This inventory will provide the roadmap for the subsequent tasks. We will review the existing documentation (e.g., Fujita et al., 2001) and compile a comprehensive table of expected sites, measured parameters, instrument vendors and models, methods, detection limits, sampling frequencies (e.g., daily, episodic), sample averaging times (e.g., 5-minute, 1-hr, 24-hr), sampling platforms (i.e., surface, upper-air, aircraft), and agency or contractor responsible for the samples. Data sets inventoried for this task also include data collected as part of CRPAQS, a large portion of which was previously inventoried by STI (Hafner et al., 2003).

The information will be summarized in tables similar to those shown in **Figure 3-1 through Figure 3-3**, which STI prepared for CRPAQS (Wittig et al., 2003). These tables provide a concise form with which to document a significant amount of information, can be easily shared among STI team members, and can be shared with future users of the data. The information will also be compiled in a database format that will be useful to subsequent tasks (i.e., can be searched, sorted, and cross-referenced).

ID	Measured Parameter	ALT1	ANGI	ANG1	ANG50	ANG95	BAC	BODB*	BTI*	COP*	EDW*	M14*	SDP	SJ4	SNFH	WAG	WGT
A	Light Scattering (integrating nephelometer with "smart" heater)	A[a]	A	A	A	A	A	A[a]	A[a]	F[a] W[a]	A[a]	W[a]	A	A	A[a]	A[a]	W
B	$PM_{2.5}$ Mass, Elements, Ammonia (Minivol with Teflon & citric acid-impregnated cellulose filters; G, XRF, AC) (Annual = 6th day)	A[a] W[a]						A[a] W[a]	A[a]	A[a] W[a]	A[a]	A[a] W[a]			A[a]		
C	$PM_{2.5}$ Ions, Carbon, Nitric Acid (Minivol with quartz & NaCl-impregnated cellulose filters; IC, AC, TOR, AA) (Annual = 6th day)							A[a] W[a]	A[a]	A[a]	A[a]	A[a] W[a]			A[a]		
D	$PM_{2.5}$ Organic Compounds (Minivol with Teflon-coated Glass Fiber; GC/MS)		A[a]				A[a]		A[a]	A[a]	A[a]	A[a]	A[a]	A[a]	A[a]		
G-1	$PM_{2.5}$ Light Absorption (black carbon surrogate) (1-wavelength Aethalometer)[c]		A[b]				A[b]	W			S	W	A[b]	A[b]		W	W
G-2	$PM_{2.5}$ Light Absorption (7-wavelength Aethalometer)[d]		F, W			W	F, W		W	F			F, W	F, W	W		
H	$PM_{2.5}$ Organic & Elemental Carbon (Thermal oxidation)		A				A										

Figure 3-1. Excerpt from a summary table showing measured parameters by site and sampling period for CRPAQS (from Wittig et al., 2003). Abbreviations are explained in the reference.

ID	Measured Parameter	Vendor/Model	Method	Averaging Time (min)	Detection Limit	Response Time (min)	Measurement Expert
A	Light scattering	Radiance Research M903 Integrating Nephelometer	Visible light scattering	5	1 Mm^{-1}	0.5	Richards
G-1	$PM_{2.5}$ black carbon (1-wavelength)	Andersen Instruments AE1X Aethalometer	Light absorption @ 880 nm	5	0.035 $\mu g/m^3$	5	Alcorn
G-2	$PM_{2.5}$ black carbon (7-wavelength)	Andersen Instruments AE3X Aethalometer	Light absorption @ 950 nm, 880 nm, 660 nm, 590 nm, 571 nm, 450 nm, 350 nm	5	0.035 $\mu g/m^3$	5	Wittig
H	$PM_{2.5}$ OC/EC carbon	Rupprecht & Patashnick 5400 OC/EC	Thermal oxidation of C to CO_2; NDIR detection	60	2 $\mu g/m^3$	60	Wittig
I-1	Particle sizing 0.3-10 µm, 16 channels	Climet Instruments Spectro.3 CI-500 OPC	Optical particle sizing and counting	5	0.0002-0.04 $\#/cm^3$ [(a)]	5	Hering
I-2	Particle sizing 0.1-2 µm, 8 channels	Particle Measuring Systems Lasair OPC	Optical particle sizing and counting	5	0.007-2 $\#/cm^3$ [(b)]	5	Hering
I-3	Particle sizing 0.01-0.4 µm, 53 channels	TSI SMPS	Scanning mobility particle sizing and counting	5	$1/cm^3$	5	Hering
J	PM_{10} mass	Met One Instruments 1020 BAM	Beta ray attenuation	60	1 $\mu g/m^3$	60	Wittig
K	$PM_{2.5}$ mass	Met One Instruments 1020 BAM	Beta ray attenuation	60	1 $\mu g/m^3$	60	Wittig
O	NO/NO_y	Thermo Environmental Instruments 42CY NO_y	Chemiluminescence with single external converter	5	100 ppt	1.3	Fitz
P	O_3	Advanced Pollution	UV absorption at 254 nm	5	1 ppb	0.2	Wittig

Figure 3-2. Excerpt from a summary table showing measurement information by parameter for CRPAQS (from Wittig et al., 2003). Abbreviations are explained in the reference.

ID	Measured Parameter	Sample Type	Samples Per Day	Sampling Times (PST)
B[a]	$PM_{2.5}$ mass, elements, ammonia (Minivol)	6th day IOP	1 1	0000-2400 0000-2400
C[a]	$PM_{2.5}$ ions, carbon, nitric acid (Minivol)	6th day IOP	1 1	0000-2400 0000-2400
D[a]	$PM_{2.5}$ organic compounds (Minivol)	6th day	1	0000-2400
L	$PM_{2.5}$ mass, elements, ammonia (SFS with aluminum denuder, Teflon filter, citric acid-impregnated cellulose filter)	Daily IOP	1 5	0000-2400 0000-0500, 0500-1000, 1000-1300, 1300-1600, 1600-2400
M	$PM_{2.5}$ ions & carbon (SFS with aluminum	Daily IOP	1 5	0000-2400 0000-0500, 0500-1000, 1000-1300, 1300-1600,

Figure 3-3. Excerpt from a summary table showing measured parameter, sampling frequency and duration information for CRPAQS (from Wittig et al., 2003). Abbreviations are explained in the reference.

3.3 TASK 2. CONFIRM EXISTENCE OF EXPECTED CCOS DATA SETS

The objective of this task is to confirm that all expected CCOS field data sets identified in Task 1 exist in the CCAQS database and, if any expected data sets are missing, to work with the CCAQS database manager to take remedial action (e.g., contacting data contractors and requesting submittal of missing data), if possible.

To do this task, we will need to screen the complete CCOS database. It will be most efficient for us to do this at STI, so we will need to transfer a copy of the database to our servers. The current size of the CCAQS database is over 120 GB and we estimate the CCOS-related data represents over half (greater than 60 GB) the size of the CCAQS database. Because of the size of the CCOS data set, we believe that to complete this task it is most efficient to obtain a full copy of the most recent CCAQS database rather than try to extract the CCOS data sets using the existing CCAQS web-based interface. We will work with the CCAQS database manager to obtain a copy of the CCAQS database in Microsoft SQL Server data file (MDF) format. We propose to provide ARB with a Windows-compatible, external disk drive (200 GB Integrated Drive Electronics [IDE] or larger drive, USB- or FireWire-compatible) on which to copy the existing CCAQS MDF file and send back to STI. The MDF file will be loaded onto STI's Microsoft SQL Server 2000 system.

This approach gives us direct access to the CCAQS database and allows us to efficiently extract and analyze selected data sets. We will use the Microsoft SQL Server tools, such as stored procedures, to automate data extraction and develop automated data checks. We may also develop automated data extraction and analysis tools using the Microsoft .NET framework. Besides being able to develop automated tools, STI data analysts will be able to directly access the SQL database and run ad hoc queries to confirm the existence of selected data sets and to check data set consistency. We will also be able to efficiently deal with both raw data and resubmitted data as they exist in the database. Also, any tools we develop for this project that work against the CCAQS database can potentially be reused by ARB in the future.

We will use a combination of automated data set checks, manual paper trail checks, and selected manual data checks to identify missing or incomplete data sets. Automated data set checks will be used to summarize overall data set metrics. For example, we will categorize each data set by sampling platform (i.e., surface, upper-air, and aircraft), site, and parameter and summarize data by sampling frequency, dates of collection, range of quality control codes, data minimum and maximum, and number of data values. Paper trail checks will include review of reporting agency and contractor documentation identified in Task 1 for information about the type and quality of data submitted and confirmation that data exist in the database as documented. Manual data checks will target selected data for specific sites, parameters, and date ranges and include visual inspection of data values, quality control flags, and comments submitted with the data.

The deliverable for this task will be an inventory of available data, an inventory of missing data, and a list of suggested remedial actions for the missing data. Using the missing data inventory, we will work with the CCAQS data manager to facilitate fixing the identified problems. Whenever possible, working with the CCAQS data manager, we will leverage existing CCAQS data ingest tools and procedures for re-submittal of data.

3.4 TASK 3. FLAG GROSS OUTLIERS IN THE DATABASE

The objective of this task is to identify and flag gross data outliers for ozone, its precursors, and meteorological data collected during the CCOS study. The purposes of undertaking this task are to (1) identify the magnitude of the problem of "bad" data in the CCAQS database, (2) for selected days, flag gross outliers that will undergo more rigorous review as part of Task 6, and (3) prepare the quality-assured data for resubmission into the CCAQS database. For this task, we define gross outliers as data that do not fit the expected physical, spatial, and temporal characteristics of the parameter.

Our approach to performing this task includes four main steps.

1. Step 1 is to define meaningful quality-control checks by relying on team members—measurement experts—who have in-depth knowledge of the meteorological and air quality data collected as part of CCOS. The organizations that collected the data will also be consulted to define these QC criteria. This approach was used successfully by STI in the collection and validation of CRPAQS data.
2. Step 2 is to use our in-house database experts and quality-control software tools to create algorithms to automatically and efficiently perform the checks and to flag gross outliers. Note that we will not overwrite existing data flags, but rather create new flags. Some of these tools were developed for CRPAQS.
3. Step 3 is to identify particularly problematic data sets, discuss remediation options with the TC, and take remedial action, if possible.
4. Step 4 is to prepare the selected data sets for resubmission into the CCAQS database.

Based on our experience with data validation, we have defined a preliminary list of quality-control checks, which are summarized in **Table 3-1**. The checks are designed to catch gross outliers by analyzing the data from spatial, temporal, chemical, and physical perspectives. These checks include range (i.e., minimum, maximum values), nearby measurement (i.e., "buddy" checks of concentrations among sites that typically behave similarly), rate of change (e.g., change in concentration from one sampling period to the next), monitor "sticking" (i.e., repeated values for more than a set number of sampling periods), low values (e.g., nighttime ozone concentrations in urban areas should be low due to titration from fresh NO emissions), and species consistency (e.g., the sum of NO + NO_2 should not be greater than NO_x at a given site and time). The checks can be automated using SQL Server queries and the output reviewed by the experts. The experts then decide which QC flags to assign to the identified data. It is likely that the focus will be to flag data as "suspect", comment on why the data were flagged (e.g., failed consistency check / $NO>NO_x$), and recommend the next validation step as part of Task 6 (e.g., review time series data, inspect instrument operating range, compare data to additional sites). These types of checks have proven to be effective in several of our current and past projects such as EPA's AIRNow program (Dye et al., 2003), the 1997 Southern California Ozone Study (MacDonald et al. 2001), and CRPAQS data management (Hafner et al., 2003). Output from Task 2 will also be useful for this task.

The outcome of this task will be a summary of data and data sets that need additional quality control, recommendations on QC actions, and delivery of re-QC'd data sets to be resubmitted for processing back into the CCAQS database.

Table 3-1. Summary of the types of automatic data checks to be applied to the CCOS data.

Check	**Description**	Parameter(s)
Range: Check of maximum values by site and sampling period.	If [parameter] > maximum or < minimum, then parameter is flagged as a gross outlier.	Ozone, ozone precursors, surface and aloft meteorology
Buddy: Compare data value to average value of surrounding stations with similar monitoring environment (urban, rural, etc.).	If [parameter–Buddy] > criteria, then parameter is flagged as suspect. If the value being checked is not within the user-specified range (typically 20-40 ppb for ozone) of the Buddy value average, the data are flagged as suspect. A minimum number of Buddy sites (typically two to four are defined) are necessary for the check to be performed.	Ozone, ozone precursors, surface meteorology
Rate of Change: Typically applied to continuous data—compares the rate of change in parameter from one hour to the next; when the difference (or change) exceeds criteria set for each hour and for each site, the data are flagged as suspect.	If [parameter(hr x) – parameter(hr x-1)] > criteria, then parameter is flagged as suspect.	Ozone, ozone precursors, surface and aloft meteorology
Sticking: Check to determine whether values remain unchanged for a specified number of sampling periods. The check can be tailored for specified time periods. For example, ozone values below X ppb (typically 40 ppb) often remain at a fixed value during the overnight hours and thus will not be checked.	If [parameter] < X, then sticking check not applied. If [parameter] > X and Y continuous hours (typically 3) occur with no change in value, then parameter value is flagged as suspect.	Ozone, ozone precursors, surface meteorology
Species Consistency: Check to determine consistency between species by checking ratios and sum of species	Checks include the following expectations (if not met, flag as suspect): $NO_x \geq NO + NO_2$ $NO_y \geq NO_x$ Total VOC ≥ sum of identified VOCs Typically abundant VOCs present above detection (e.g., toluene, ethane, i-pentane)	Ozone precursors

3.5 TASK 4. CHECK CCOS DATA SETS FOR CONSISTENCY

The objective of this task is to ensure that we have a thorough understanding of the CCOS "metadata". Metadata are the supporting information that describes the data value. In Task 2, checks will have been made of the data values themselves; this task focuses on the other information in the database. We will extract, summarize, and review for reasonableness other information from the database including time stamps (e.g., begin and end times, time standard); duplicate data records; units by parameter (e.g., are there multiple units for a given parameter?); station coordinates (e.g., are the coordinates correct?); data labels (e.g., are canister labels unique to samples?); and proper labeling of collocated, duplicate, and replicate samples. Because we are dealing with a large database, we will develop a simple automated tool to extract and summarize data wherever possible using the MS SQL Server tools described in Task 2. The results will then be reviewed and summarized.

Data gaps will be documented by site and parameter in a manner similar to that illustrated in **Figure 3-4** as prepared by Wittig et al. (2003) for CRPAQS data. As much as possible, we will automate the extraction, summarization, and display of data gap information for efficiency. These figures will allow modelers and analysts to quickly understand the availability (and, when updated after later tasks, the quality) of data collected during selected time periods during CCOS.

3.6 TASK 5. AUDIT THE QC PERFORMED ON THE CCOS DATA

The objective of this task is to summarize the level of QA/QC of the CCOS data sets performed by data contractors and subsequent data users. A thorough understanding of the quality of the data, what and how QA/QC steps have been applied, and how the validation results have been documented is vital to data analysts and modelers who need to use the data. This task consists of two parts, as described below.

3.6.1 Identify Data Validation Milestones

The first part of this task is to identify major data validation milestones such as what QC steps were applied to the various data sets, identify the QC codes used by each agency/contractor validation specialist, and identify the date of submittal of each validated data set by the reporting agencies and contractors. The CCOS field study documentation laid out the QC level definitions for data submittals to follow. All reporting agencies and contractors were to perform Level 0 validation prior to data submittal (raw data). Level I validation was also to have occurred, but these data sets *may not* have been resubmitted to the CCAQS database as required. Additional QC—Levels II and III—may have been performed by data analysts and modelers and these data sets have *likely not* been resubmitted to CCAQS.

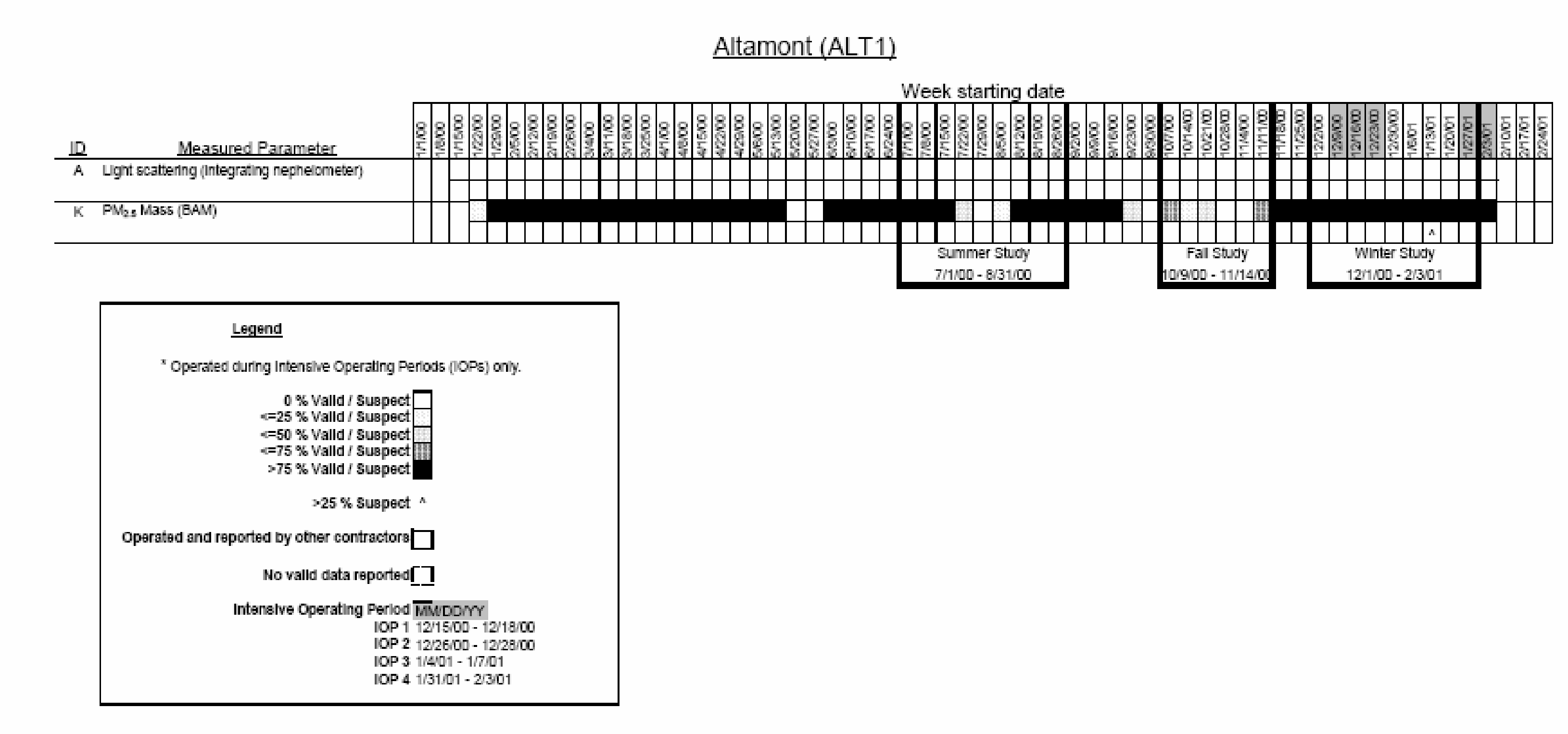

Figure 3-4. Example data availability and validation summary from CRPAQS measurements (Wittig et al., 2003).

Ensuring that Level I, II, or III validated data actually reside in the database is key to this part of the task. The QC codes, data validation level, and date of submittal will be obtained from the database. The QC steps carried out on the data will need to be obtained through documentation from the staff that validated the data including DRI, ARB, the air quality management districts (AQMDs), and other contractors. We will work with the CCOS TC to identify the appropriate contacts and to obtain validation documentation. We will review available documentation from these agencies regarding their data validation efforts and findings and summarize this information for other users. If needed, we will contact these agencies for more information and discussion.

3.6.2 Research Previously Identified Data Issues

The second step in this task is to interview selected data users (e.g., analysts and modelers) who have identified issues in the data. Analysts and modelers subject the data to additional scrutiny beyond the initial QA/QC that can result in discovery of more data that need to be invalidated, changed, or flagged as suspect. This ongoing process is consistent with Level II and III validation that occurs with all large data sets. We will work with the TC to identify these users; likely candidates include John DaMassa or Ajith Kaduwela at ARB; Mike Kleeman at University of California (U.C.) Davis; Rob Harley at U.C. Berkeley; Gail Tonnesen at U.C. Riverside; Eric Fujita, Bill Stockwell and Bob Keislar at DRI; Don Lehrman at T&B Systems; and Charlie Blanchard at ENVAIR. We will then develop a questionnaire to send to the identified analysts and modelers. The questionnaire will facilitate, clarify, and help to document our subsequent discussions with the identified analysts and modelers. Questions will include the following:

- *Were the data used obtained from the CCAQS database or through another mechanism?* It is possible that data were obtained directly from the reporting agency/contractor (although this was not encouraged or desirable).
- *What data were obtained from the CCAQS database?* We will need to know which sites, dates, and parameters were pulled from the database.
- *What was the date of the data extraction from the CCAQS database?* We will need this to compare to our understanding of the data set validation level and submittal obtained earlier in this task.
- *Were the data extracted through the web site by the analyst/modeler or were the data provided by ARB?* In some cases, the analyst/modeler may have worked with ARB to extract a large amount of data without going through the web site.
- *What was the QC level of the data?* The data should have a QC level code and possibly notes associated with the data values to give some indication of level.
- *How were the data intended to be used by the analyst or modeler?* Types of QC efforts and levels of data "cleaning" needs are associated with different types of analyses. For example, some types of data analyses are sensitive to outliers and thus more data might be identified as suspect or invalid for that analysis relative to an analysis that uses central tendencies in the data or long-term averages.

- *What problems or issues were identified in the data?* For example, when the data were compared among sites or among samplers were there gross outliers, biases, or abrupt changes in concentration noted? Sometimes these potential problems are not noted when the reporting agency/contractor validates only their own data.
- *How were the questionable data documented in the data set used?* For example, we need to know if QC flags were used that are different from those currently in the CCOS database.
- *How were the data treated (as invalid, suspect) in the analysis or model effort?* This will be important to the wider applicability of the validation decisions made by the analyst/modeler to other users of the data. One analyst's suspect data may be useful to another analyst.
- *What remedies are recommended by the analyst/modeler?* For example, do the data need to be adjusted for a bias?

The deliverable for this task will be a summary of our review of major data validation milestones and our interviews of analysts and modelers. We will also provide recommendations of potential next steps.

3.7 TASK 6. DETAILED QUALITY ASSURANCE/QUALITY CONTROL

The objective of this task is to ensure that reliable meteorological and air quality data sets that are at a consistent level of QA/QC are ready for immediate use for data analysts and modelers without need for further judgment regarding data quality. Because there is a need by ARB and other stakeholders for a portion of these data sets to be available early in the project schedule, some QA/QC will be conducted in parallel to Tasks 1 through 5. The "fast-track" high-priority data sets will be defined by the CCOS TC at the beginning of the project. Our current understanding is that the TC will work with the contractor to identify several high-priority data sets that are important for photochemical modeling.

As a result of findings from Tasks 1 through 5, and upon consultation with the TC, we will identify, prioritize, and then QA/QC other selected (i.e., non-fast-track) data sets. Considerations for additional high-priority data sets include periods that are representative of different ozone event types, intensive operation periods when additional data are available including aloft air quality data collected by aircraft (these data are important for the verification and evaluation of above-ground model results), parameters important to ozone, and periods with good data recovery and quality (as determined from prior tasks).

For both fast-track and non-fast-track data validation, we have assumed that the focus will be on 1-hr average data for continuous air quality and meteorological measurements, 24-hr averages for the VOC canisters, and various averaging times for the aircraft measurements. We will work with the TC to select priority data sets with which to explore more detailed temporal validation (e.g., 5-minute averages) based on analysis/modeling needs and available funds after validation of the data sets of principal focus.

The result of the QA/QC effort will be a new set of data flags applied to the data sets that will either be the same as the existing data flags or will be updated to reflect a more appropriate validation based on our QA/QC. We assume that the data for this QA/QC task will be available from the CCAQS database or be made available by contractors in the CCAQS submittal format.

Our overall approach to meet the objective of this task is for data experts to lead the quality assurance efforts. STI's data experts are supported by experienced staff and the Principal Investigator. These experts have written guidance documents for, and have led several training sessions on, the quality assurance of the data types collected for CCOS (Lindsey et al., 1995; MacDonald et al., 2001; Main et al., 1996; Main and Prouty, 2000; Hafner et al., 2004; Hafner, 2003). STI will rely on both custom (e.g. VOCDat and SurfDat) and commercial (e.g. SYSTAT, Igor, and GraphXM) programs to efficiently QC the data.

The following subsections provide descriptions of procedures for QA/QC of the types of data collected under CCOS sponsorship. We will apply appropriate procedures to the data sets identified by the TC as high priority. The precise amount of time and resources (and, therefore, the magnitude of data to be QA/QC'd) for both the fast-track and non-fast-track data sets will be determined in conjunction with the TC.

3.7.1 Surface Air Quality Data

For the QC of the surface air quality data, STI will use in-house data visualization software developed to facilitate graphical review of selected data sets. The software allows the reviewer to plot several parameters at a time (e.g., ozone, NO, and NO_y; selected target VOCs); to change QC flags; and to annotate changes to QC flags. We will also use screening criteria developed through our experience with aerometric data to identify potentially suspect or invalid data. The gross outliers (such as sum of targeted VOC species > total VOC) will have been identified in an earlier task—this task will focus on the details (e.g., expected relationships among species).

Data validation for continuous monitors (such as ozone, NO_x, NO_y) will include graphically reviewing time series plots of pollutant concentrations paying particular attention to times before, during, and after calibrations, maintenance, and other off-line periods. We will inspect data spikes, dips, and outliers. This process is facilitated by plotting complementary data together (e.g., ozone and NO/NO_y).

With the help of measurement experts, screening criteria were prepared during CRPAQS for ozone, NO, and NO_y to assist in data validation. These criteria are listed in **Table 3-2**.

Table 3-2. Example screening criteria.

Instrument	Parameter	Screening Criteria
Ozone	Ozone	Should not drop below -2 or exceed 200 ppb
		Point-to-point variation should not exceed 30 ppb
		Six consecutive values should not be equal. *This test was not applied between 0000 and 0600 hours.*
NO/NO_y	NO/NO_y urban	Should not drop below -1 nor exceed 700 ppb
		Point-to-point variation should not exceed 50 ppb
		NO should not exceed NO_y
		30 consecutive values should not be equal
NO/NO_y	NO/NO_y rural	Should not drop below -1 or exceed 300 ppb
		Point-to-point variation should not exceed 30 ppb
		NO should not exceed NO_y
		30 consecutive values should not be equal

Some of the PAMS hydrocarbon and carbonyl compound data presumably reported to the CCAQS database (i.e., the San Joaquin Valley PAMS sites) are currently being validated by STI under separate contract and the validated data will be available within the timeframe set forth in this proposal. Data validation steps include the preparation and review of summary statistics, application of screening criteria, inspection of time series plots of all target species, and inspection of scatter plots and fingerprint plots to investigate internal consistency. The analyst starts with the total nonmethane organic compound (TNMOC) mass and species groups (i.e., paraffins, olefins, aromatics, unidentified, sum of PAMS target species, and carbonyl compounds) and then inspects each target specie. The analyst also looks at the fingerprint of every sample. STI's VOCDat software facilitates the efficient and quick performance of these steps. To perform the QC of the surface VOC measurements, we recommend the following procedures:

- Run VOCDat's auto-QC checks and inspect samples identified as failing screening criteria. The screening checks include checking whether abundant species concentrations are above a threshold (usually of 0.5 to 1 ppbC), comparing species concentrations with expected relationships (e.g., o-xylene < m-&p-xylenes), and identifying outliers (e.g., concentrations more than 3 standard deviations from the mean).
- Inspect time series plots of every specie, species group, and the TNMOC.
- Prepare scatter plots of benzene/acetylene, benzene/toluene, i-pentane/n-pentane, i-butane/n-butane, 2-methylpentane/3-methylpentane, m-&p-xylenes/o-xylene, decane/undecane, ethane/propane, and other species combinations specific to problems observed at a site.
- Inspect fingerprints. These plots help the analyst inspect patterns within samples and among samples and identify missing species data.

3.7.2 Surface and Upper-air Meteorological Data

For the QC of the meteorological data, STI recommends performing Level 1 and Level 2 validation of the RWP and SODAR winds, RASS virtual temperature, and CCOS-sponsored surface and rawinsonde meteorological data. These validation steps are a subjective review of the data that includes checks for internal (Level 1) and external (Level 2) consistency and reasonableness for each individual site for each hour.

STI will rely on experienced staff to recognize and identify common problems in each data set associated with the following types of issues:

- RWP wind data. Interference from migrating birds or precipitation, ground clutter, velocity folding, errors associated with the processing method, and instrument setup.
- RASS T_v data. Inappropriate temperature range setting, radio interference, cold bias, and inaccurate measures of vertical velocity and instrument setup.
- Sodar wind data. Fixed echoes (ground clutter) and other noise interference, and instrument setup.
- Surface meteorological data. Incorrect cross-arm directions for wind sensors, relative humidity measurements above 100%, solar radiation measurements greater than 0 at night, and other instrument setup problems.

For the Level 2 validation of the meteorological data, an experienced meteorologist will subjectively review the data for external consistency and reasonableness by comparing collocated and nearby measurements. For the surface data, this will include the creation and review of spatial plots of hourly data. For the upper-air data, the reviewer will rely on other meteorological data such as Eta Data Assimilation System (EDAS) data and National Weather Service (NWS) upper-air charts to evaluate the spatial consistency of the winds and other meteorological parameters based on the large-scale meteorological patterns.

3.7.3 Aircraft, Ozonesonde, and Lidar Measurements

STI will follow similar criteria when QC'ing the aircraft, ozonesonde, and ozone LIDAR measurements as for the QC of the surface air quality data. However, we recognize that large variations in pollutant concentrations can occur over short vertical and horizontal distances as the aircraft moves between different layers or as ozonesondes ascend, and will consider this spatial variation in the validation process. For this task, we understand that on any given day there may be measurements from five aircraft (i.e., U.C. Davis, Pacific Northwest Laboratory, Tennessee Valley Authority, and two STI-operated aircraft). We assume that we will QC the following aircraft measurements that are important to ozone modeling: ozone, NO, NO_y, temperature, humidity, winds, VOCs, carbonyl compounds, and PAN/NO_2. To perform the QC of these aloft measurements, we recommend the procedures below, which include flagging suspect data. Note that STI has performed similar procedures on the aircraft measurements taken by STI during the CCOS summer field study.

- Create spatial plots of the continuous aircraft measurements and surface pollutant concentrations at nearby times and locations for all flights for the selected episodes.

- Review the aircraft plots by (1) comparing the measurements among aircraft when they were at similar locations and times, and (2) comparing the aircraft measurements to nearby surface concentrations when the aircraft was within the mixed layer.
- Create vertical profile plots of the ozonesonde measurements and aircraft vertical spiral measurements, and time-height cross-section plots of LIDAR ozone measurements. Compare the vertical measurements from the various platforms when they are nearby in space and time. Verify that the vertical structure is consistent with the diurnal evolution of the boundary layer and the diurnal evolution of pollutants. For example, in the early morning hours in urban areas, we expect sharp pollutant gradients between the nocturnal boundary layer and the residual layer. Another example is that when NO concentrations are high, ozone concentrations should be low and visa-versa.
- For the VOC canister and carbonyl compound measurements, perform the same types of screening and graphical inspections of the data as for the surface data. However, we also recommend plotting the data as a function of altitude with particular attention to whether the sample was collected above or within the mixed layer. Previous experience with these data have shown that samples collected above the mixed layer appear more aged relative to samples collected within the mixed layer. Aging is expressed, for example, by the relative removal of more reactive compounds (i.e., xylenes) to less reactive compounds (i.e., benzene).

The output of this level of validation on all the selected data sets will be documented, and validated databases will be ready for resubmittal to the CCAQS database and for subsequent use by data analysts and modelers.

3.8 TASK 7. FINAL REPORT

The objective of this task is to summarize the results and findings regarding the quality and availability of CCOS data in the CCAQS database. The report must provide a comprehensive description of the quality of data, the specific steps that were taken to ensure adequate data quality, a summary of issues and problems that were encountered during the course of this project, and the recommended remedies (if the remedy was not applied). To summarize data availability, we will compile tables similar to those prepared for STI's CRPAQS data submittal as shown in **Figure 3-5**. We will also include a discussion about our synthesis of how the data availability and quality may affect subsequent analyses and modeling efforts.

Monitoring Site	Total No. of Records	No. of Expected Records	Percent Capture[a]	No. of Valid Records	Percent Recovery[b]	No. of Suspect Records	No. of Invalid Records	No. of Missing Records
Angiola Trailer (5-min)	114,081	114,081	100%	99,238	87%	3,574	8,272	2,997
Angiola Trailer (60-min)	9,507	9,507	100%	8,475	89%	308	503	221
Angiola 100-m Tower (5-min)	19,030	19,030	100%	18,152	95%	2	186	690
Angiola 100-m Tower (60-min)	1,586	1,586	100%	1,513	95%	1	19	53
Sierra Nevada Foothills (5-min)	21,489	21,489	100%	20,321	95%	6	1132	30
Sierra Nevada Foothills (60-min)	1,791	1,791	100%	1,738	97%	2	51	0

[a] % capture = total number of records/expected records*100%.
[b] % recovery = number of valid records/total number of records.

Figure 3-5. Example data summary for ozone from STI's CRPAQS data submittal (Hyslop et al., 2003).